Journal of Informatics and Advanced Data Science 0(0)

Efficient Industrial Casting Inspection using Deep Transfer Learning

Qusay Shihab Hamad [1], Hilal A. Fadhil [2], Jaafar A. Aldhaibani [1], Mazen Abdullatif [3,4], Mazin S. Al-Hakeem [5], Sahar Saleh [6,7], Tale Saeidi [6]

[1] University of Information Technology and Communications (UoITC), Baghdad, Iraq
[2] Faculty of Engineering, Electrical and Computer Eng., Sohar University, Sohar, Oman
[3] Cordoba University, Aleppo, Syria
[4] AL Esraa University, Baghdad, Iraq
[5] Al-Farabi University, Baghdad, Iraq.
[6] WiSAR Lab, Atlantic Technological University, Letterkenny, Ireland.
[7] University of Aden, Aden, Yemen

*Corresponding Author: qusay.phd@gmail.com

Abstract

Aluminum alloy castings have become vital elements of numerous contemporary sectors and their structural integrity directly influences the mechanical safety. It is due to this that quality control throughout the manufacturing process is of importance. In this paper, it is proposed to use deep learning to identify the defects in cast products using digital images. A transfer learning approach with a VGG-19 architecture was used to address the difficulties of large-data demands and sophisticated tuning of hyperparameters. The model was also tested on a large data set of 8,012 images including normal and defective samples. We have found that the model has a test accuracy of 99.251 and a recall of 99.265. These results indicate that the presented transfer learning model is a very accurate and robust one which can provide a practical solution to the quality control in real-time within the industrial setting.

Keywords

Industry 4.0, Transfer Learning, Casting Defect Detection, VGG-19, Automated Quality Inspection.

1. Introduction

Casting aluminum alloy is a giant component of contemporary production, and it is found everywhere, including automobile engines or even airplane components. Since these components usually need to work with significant amounts of stress, their quality becomes a significant safety issue. When a small internal crack is present in a part or a pore that will not be noticed by anyone, it may result in complete mechanical failure in the future. Factories have long been using people to examine each part or X-ray image. However, as production lines become faster it is not realistic to ask humans to remain 100 per cent focussed hours. Individuals become weary and various inspectors may hold varying views on what constitutes a defect.

Due to these human constraints, the trend of more companies shifting to Industry 4.0 and smart automation is growing. According to recent studies published in 2025 and 2026, deep learning has become the tool of this type of work. As an illustration, [12] has just indicated that we have numerous new algorithms, but the difficulty lies in getting them fast and reliable enough to operate on an actual factory floor. Other authors such as [13] have devoted particular attention to the small object problem in which he stated that a small gas bubble in a large casting is extremely difficult to detect by an ordinary computer..

Acquiring sufficient data is one of the largest of headaches in this field. You could get one 1000 parts of a high-quality factory to every 1 broken parts. It implies that it is extremely difficult to gather thousands of defect images to train a new AI model and start with a clean slate. It is this lack of data that has made transfer learning so popular in the recent past. With a model such as VGG-19, which already has a sense of what shapes and textures should look like, it is just necessary to train it on the exact patterns of casting flaws. As demonstrated in [14], this method can maintain accuracy of over 98 percent even with 40 percent less data than normal.

Our analysis is no exception. We decided to work with a pre-trained VGG-19 model since it is a workhorse-type architecture that is stable and has a good balance between high accuracy and the speed required in an industry. With the goal of processing a large number of images (more than 8000) we want to demonstrate that you do not require the most complicated or latest model to achieve almost perfect results, that a properly-configured classic model used with transfer learning can also be a better engineering decision.

The key points made in this paper are: (1) We present a model which relies on a pre-trained CNN (VGG-19) to predict defective industrial casting images. The suggested model can automatically identify features based on image-level annotations and train the attention maps of the castings. (2) 8,012 images form a large dataset with which we assess and test the effectiveness of the proposed model. The findings indicate that the model suggested is very good in performance.

The remaining paper is organized in the following manner: Section 2 presents related work, whereas Section 3 describes our methodology. In Section 4, we present the results and leave our conclusive ideas in Section 5.

2. Related Work

Industrial quality control was one of the areas where deep learning has made a significant transition as factories attempt to abandon slow manual inspection. Initial studies on this field were concerned with simple machine vision. Indicatively, one of the earliest to demonstrate how to use X-ray images to locate casting defects in metals such as aluminum and titanium was Ji et al. [1] and Du et al. [2]. More recently, Gupta et al. [3] showed that, with the help of ensemble models such as

ResNet50, one can attain over 98% accuracy on large-scale casting datasets, which means that deep learning is prepared to be used on high volumes in industries. With the enhancement of technology, scientists such as Wang et al. [4] and Pastor-Lopez et al. [5] have created systems which could identify fine surface defects despite the fact that the dataset available was relatively small.

Researchers have provided specialized methods to increase the speed and detail of detection. Wu et al. [6] and Jiang et al. [7] employed data augmentation and dense networks to enhance the feature extraction of X-ray pictures.. Others such as Han et al. [8] were interested in the segmentation of defects in specialized materials such as silicon wafers. Moreover, Wang and Jing [9] pointed out that fine surface defects like burrs on cast iron should be captured with matters of detail and better feature pyramids of various scales of defects.

Nonetheless, one of the biggest challenges that are still there is the data bottleneck the challenge of gathering thousands of pictures of uncommon defects in a quality production set up. This is what makes transfer learning the choice of strategy in the recent researches. Wang and Yu [10] observed that pre-trained weights greatly enhance the accuracy of the model over training-from-scratch as the model will have a more accurate idea of visual textures. This is endorsed by Tanyildiz and Şahin [11], who projected that transfer learning will assist manufacturers to avoid the cumbersome procedure of educating the new models on Industry 4.0 settings. Likewise, Song et al. [12] have just conducted a review of the state of the art and observed that today the tendency is to optimize the existing frameworks to be used in real-time instead of merely creating bigger and more complicated models.

This is the practical course of our work. Although a few papers, including Zhang et al. [13], suggest extremely intricate "bidirectional" extraction techniques of minute flaws, numerous industrial applications will be served better by stable and reliable architectures. As an example, Lin et al. [14] showed that a carefully-tuned trained model can retain a high accuracy of more than 98% even after training data is reduced by 40 percent. Our study will present a lightweight and highly accurate solution by selecting VGG-19 and changing the last classification layers to achieve the real demand of a modern casting facility.

3. Methodology

The goal of this research is to create a reliable system that can automatically identify defects in aluminum casting images using deep learning. Instead of building a new model from scratch, we used a transfer learning strategy to leverage the power of a well-established architecture.

3.1 Dataset Description

To train and test our system, we used a large dataset containing 8,012 images of casting products. The dataset is divided into two main categories: 4,613 samples showing various defects and 3,399 samples of normal, healthy parts. Fig. 1 shows a few examples of these images. To ensure the model could generalize well to new data, we split the images into 60% for training and 40% for testing. This large volume of data is important because it provides enough variety for the model to learn the different textures and shapes associated with casting flaws.

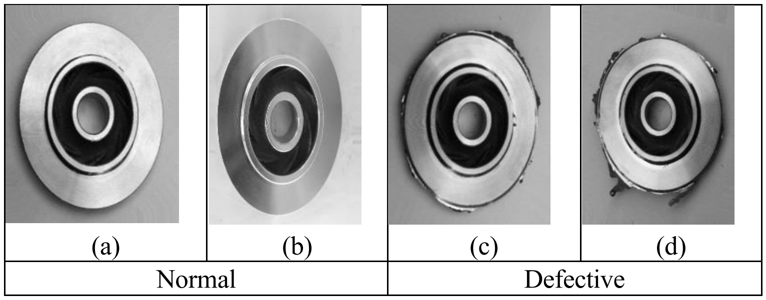

Fig. 1. Various normal and defective casting case images

3.2 Transfer Learning Strategy with VGG-19

We chose the VGG-19 architecture as the core of our system. While there are many newer and more complex models available in 2026, VGG-19 remains a workhorse in the industry because it is stable and very effective at extracting high-level features from images. The basic structure of this architecture is illustrated in Fig. 2. By using transfer learning, we took the weights that the model had already learned from millions of general images and applied them to our casting problem. This is a huge advantage because it allows the model to achieve near-perfect results without needing the massive computing power that training from scratch would require. In a real factory setting, this efficiency is often more important than having the most complex algorithm.

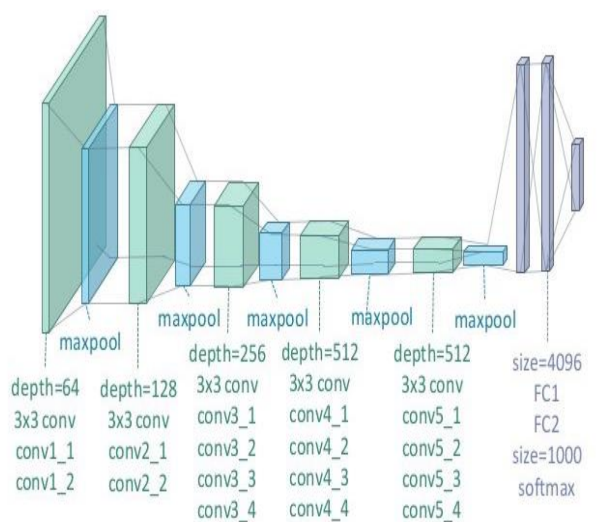

Fig. 2. VGG-19 architecture.

3.3 Model Architecture and Modification

Our proposed model keeps all the standard convolutional and max-pooling blocks of the original VGG-19 to handle feature extraction. However, we made a specific change to the final part of the network to fit our two-class problem (Defective vs. Normal). This modified structure, which we use for our detection process, is shown in Fig. 3.

- **Layer Replacement:** We replaced the standard Softmax classifier, which was originally designed for 1,000 different categories, with a new version specifically for our two classes.
- **Classification:** The model generates a similarity score. If the score is higher than 0.5, the part is classified as normal; otherwise, it is flagged as defective.
- **Automatic Feature Extraction:** The system is designed to learn directly from image-level annotations, which means it can automatically find and "pay attention" to the important parts of the casting image without needing manual guidance.

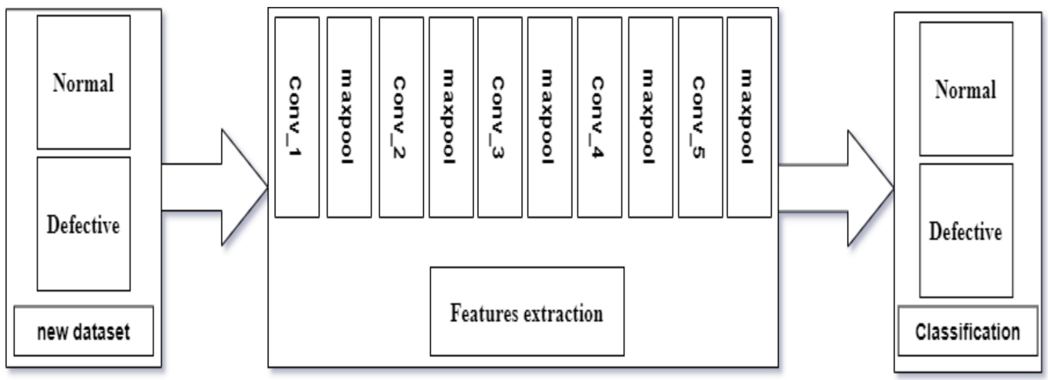

Fig. 3. The proposed model

3.4 Experimental Setup

All experiments were conducted using a system equipped with an Intel Core i5-9400F CPU and an Nvidia GeForce RTX 2070 Super GPU. We used Python 3.8 as the primary programming language. The model was trained over 100 epochs to ensure the weights were fully optimized, and we used standard performance metrics like Accuracy, Recall, and F1-score to evaluate how well the system performs in a real-world scenario.

4. Results and Discussion

The specifications of the machine utilized in this study are listed in Table 1 below. The results of the proposed model are presented in this section. Table 2 summarizes the results and compares the proposed method to state-of-the-art deep learning models. All analyses were carried out over a period of 100 epochs. Additionally, the confusion matrix was computed, as shown in Fig. 4. As demonstrated by the results, the proposed model correctly classified 3205 instances with only 24 missing instances. This study generates two output classes: 'yes', which indicates normal, and 'no,' which indicates defective.

Accuracy, precision, specificity, sensitivity (Recall), and F-score are the five metrics used to evaluate models [15][16]. These metrics were calculated using the following equations: Eq. (1) to Eq. (4).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\% \qquad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \text{ x } 100\% \qquad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \text{ x } 100\% \qquad (3)$$

$$\text{F1 Score} = \frac{2TP}{2TP+FP+FN} \text{ x } 100\% \qquad (4)$$

TABLE 1. THE DETAILED SETTINGS OF THE UTILIZED SYSTEM.

Name	Setting
Hardware	
CPU	Intel(R) Core (TM) i5-9400F
Frequency	2.90 GHz
RAM	16 GB
GPU	Nvidia GeForce RTX 2070 Super
SSD	476 GB
Hard drive	1 TB
Software	
Operating system	Windows 10
Language	Python 3.8

TABLE 2. DEFECTIVE CASTING RECOGNITION RESULTS.

Reference	Method	Accuracy (%)	Precision (%)	Recall (%)	F_1 score (%)
Lin et al.[17]	CNN from scratch	96	-	-	-
Nguyen et al.[18]	CNN from scratch	98	-	-	-
Proposed model	VGG-19	99.251	98.974	99.265	99.119

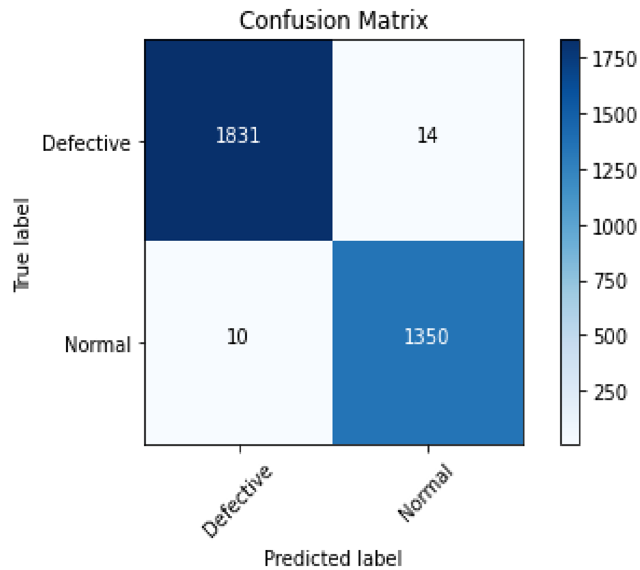

Fig.4. Confusion matrix

To check how the VGG-19 model learned during the process, we looked at the accuracy and loss curves for 100 epochs (Fig. 5 and Fig. 6).

As shown in the Model Accuracy graph (Fig. 5), the model learned very fast. It reached more than 90% accuracy in only the first 10 epochs. There are some small jumps in the results at the beginning, like around epoch 7. These small changes happen because the model is adjusting the weights in its final layers during the transfer learning phase. After that, the performance becomes very steady. A very important point is that the validation accuracy is almost the same as the training accuracy. This proves that our model works well on new images and does not have an overfitting problem.

At the same time, the Model Loss graph (Fig. 6) shows a quick drop and stays near zero as training continues. Because the gap between training loss and validation loss is very small, we can say the settings for our 8,012-image dataset were chosen correctly. These graphs support our confusion matrix results (Fig. 4), which showed that the model correctly identified 3,181 cases with very few errors. In Table 2, we compare our work to other models. Our VGG-19 method reached a test accuracy of 99.251%, which is better than previous methods.

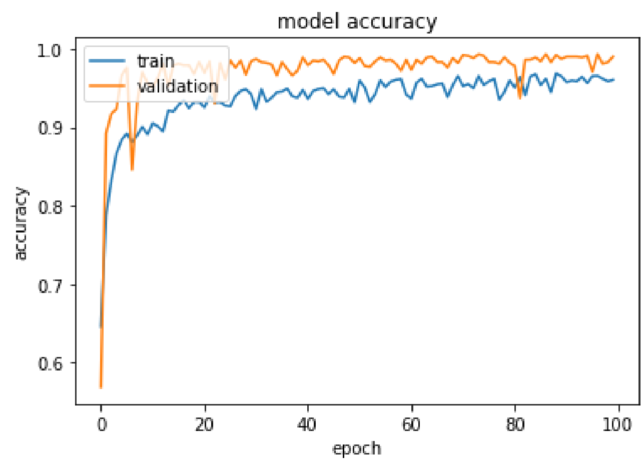

Fig. 5. Training and validation accuracy of the VGG-19 model over 100 epochs.

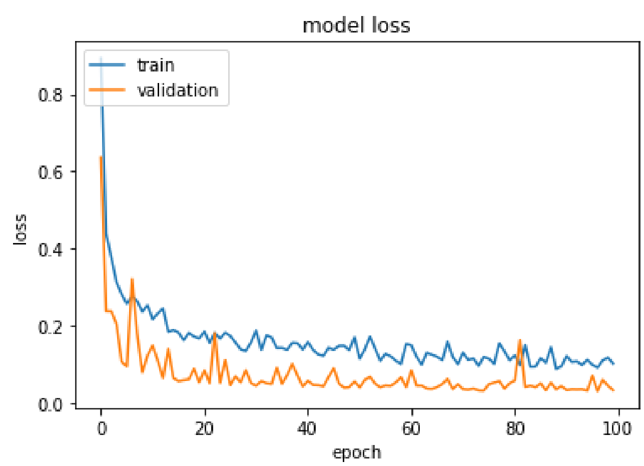

Fig. 6. Training and validation loss of the VGG-19 model over 100 epochs.

5. Conclusion

This study successfully applied a transfer learning approach using the VGG-19 architecture to automate the detection of industrial casting defects. By keeping the robust feature extraction layers of the standard VGG-19 and modifying only the final layer for binary classification, we were able to achieve superior results in terms of accuracy, recall, and F1-score. Our findings prove that using a pre-trained model is a highly efficient strategy for the manufacturing industry; it significantly speeds up the learning process and reduces the massive data requirements typical of training a CNN from scratch. Ultimately, this model provides a practical, high-performance solution for real-time quality inspection, ensuring that safety-critical components meet industrial standards without the need for exhaustive manual oversight.

References

[1] X. Ji *et al.*, “Filtered selective search and evenly distributed convolutional neural networks for casting defects recognition,” *J. Mater. Process. Technol.*, vol. 292, p. 117064, Jun. 2021, doi: 10.1016/j.jmatprotec.2021.117064.

[2] W. Du, H. Shen, J. Fu, G. Zhang, and Q. He, “Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning,” *NDT E Int.*, vol. 107, p. 102144, Oct. 2019, doi: 10.1016/j.ndteint.2019.102144.

[3] R. Gupta, V. Anand, S. Gupta, and D. Koundal, “Deep learning model for defect analysis in industry using casting images,” *Expert Syst. Appl.*, vol. 232, no. June, p. 120758, 2023, doi: 10.1016/j.eswa.2023.120758.

[4] Y. Wang, C. Hu, K. Chen, and Z. Yin, “Self-attention guided model for defect detection of aluminium alloy casting on X-ray image,” *Comput. Electr. Eng.*, vol. 88, p. 106821, Dec. 2020, doi: 10.1016/j.compeleceng.2020.106821.

[5] I. Pastor-López, B. Sanz, A. Tellaeche, G. Psaila, J. G. de la Puerta, and P. G. Bringas, “Quality assessment methodology based on machine learning with small datasets: Industrial castings defects,” *Neurocomputing*, vol. 456, pp. 622–628, Oct. 2021, doi: 10.1016/j.neucom.2020.08.094.

[6] B. Wu *et al.*, “An ameliorated deep dense convolutional neural network for accurate recognition of casting defects in X-ray images,” *Knowledge-Based Syst.*, vol. 226, p. 107096, Aug. 2021, doi: 10.1016/j.knosys.2021.107096.

[7] L. Jiang, Y. Wang, Z. Tang, Y. Miao, and S. Chen, “Casting defect detection in X-ray images using convolutional neural networks and attention-guided data augmentation,” *Measurement*, vol. 170, p. 108736, Jan. 2021, doi: 10.1016/j.measurement.2020.108736.

[8] H. Han, C. Gao, Y. Zhao, S. Liao, L. Tang, and X. Li, “Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks,” *Pattern Recognit. Lett.*, vol. 130, pp. 234–241, Feb. 2020, doi: 10.1016/j.patrec.2018.12.013.

[9] “IET Image Processing - 2023 - Wang - Deep learning-based methods for detecting defects in cast iron parts and surfaces.pdf.” .

[10] X. Wang and X. Yu, “NDT and E International Understanding the effect of transfer learning on the automatic welding defect detection,” *NDT E Int.*, vol. 134, no. May 2022, p. 102784, 2023, doi: 10.1016/j.ndteint.2022.102784.

[11] H. Tanyildiz, “Transfer Learning for Detection of Casting Defects Model In Scope of Industrial 4 . 0 Endüstri 4 . 0 Kapsamında Döküm Hatalarının Tespiti İçin Transfer Öğrenme Modeli,” vol. 12, no. 3, pp. 45–51, 2023.

[12] Z. Song, Y. Yao, and J. Yang, “Casting Defect Detection and Recognition Technology Based,” 2025.

[13] S. Zhang, H. Li, P. Ren, T. Peng, and X. Meng, “OPEN A detection method for small casting defects based on bidirectional feature extraction,” pp. 1–18, 2025.

[14] H. Lin, H. Wu, and C. Lin, “A Deep Transfer Learning-Based Visual Inspection System for Assembly Defects in Similar Types of Manual Tool Products,” 2025.

[15] M. Toğaçar, B. Ergen, and Z. Cömert, “COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches,” *Comput. Biol. Med.*, vol. 121, p. 103805, Jun. 2020, doi: 10.1016/j.compbiomed.2020.103805.

[16] W. Jin, S. Dong, C. Dong, and X. Ye, “Hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph,” *Comput. Biol. Med.*, vol. 131, p. 104252, Apr. 2021, doi: 10.1016/j.compbiomed.2021.104252.

[17] J. Lin, Y. Yao, L. Ma, and Y. Wang, “Detection of a casting defect tracked by deep convolution neural network,” *Int. J. Adv. Manuf. Technol.*, vol. 97, no. 1–4, pp. 573–581, Jul. 2018, doi: 10.1007/s00170-018-1894-0.

[18] T. P. Nguyen, S. Choi, S.-J. Park, S. H. Park, and J. Yoon, “Inspecting Method for Defective Casting Products with Convolutional Neural Network (CNN),” *Int. J. Precis. Eng. Manuf. Technol.*, vol. 8, no. 2, pp. 583–594, Mar. 2021, doi: 10.1007/s40684-020-00197-4.